4.8 A 32nm 3.1 Billion Transistor 12-Wide-Issue Itanium® Processor for Mission-Critical Servers

Reid J. Riedlinger[1], Rohit Bhatia[1], Larry Biro[2], Bill Bowhill[2], Eric Fetzer[1], Paul Gronowski[2], Tom Grutkowski[1]

[1]Intel, Fort Collins, CO,
[2]Intel, Hudson, MA

The next generation in the Intel® Itanium® processor family, code named Poulson, has eight multi-threaded 64 bit cores. Poulson is socket compatible with the current Intel® Itanium® Processor 9300 series (Tukwila) [1]. The new design integrates a ring-based system interface derived from portions of previous Xeon® and Itanium® processors, and includes 32MB of Last Level Cache (LLC). The processor is designed in Intel®'s 32nm CMOS technology utilizing high-K dielectric metal gate transistors [2] combined with nine layers of copper interconnect. The 18.2×29.9mm^2 die contains 3.1 billion transistors, with 720 million allocated to the eight cores (Fig. 4.8.1). A total of 54MB of on die cache is distributed throughout the core and system interface. Poulson implements twice as many cores as Tukwila while lowering the thermal design power (TDP) by 15W to 170W and increases the top frequency of the I/O and memory interfaces by 50% to 6.4GT/s.

The design introduces a new core micro-architecture and floor plan that significantly improves frequency, and power efficiency. The core implements an 11-stage in-order, decoupled frontend and backend pipeline which employs replay and flush mechanisms versus the previous global stall micro-architecture. The decoupled pipelines enable an increase in resource utilization and throughput. The frontend pipeline fetches six instructions per cycle while the backend executes and retires up to twelve instructions per cycle. The backend execution resources include six ALUs, two integer units, two floating-point units, two memory units and three branch units distributed across twelve ports. A 96-entry distributed instruction buffer, replicated for each thread, decouples the frontend and backend pipelines while storing replay information. To further reduce data access penalties the core implements new features including: a hardware data prefetcher, data access hints and improved concurrent TLB accesses. The core improves performance through fine grain multi-threading, replicated instruction buffers and D-side TLBs, a 4 cycle integer multiplier, and an additional 32 entries to the integer register file. A three-level cache hierarchy supports these parallel execution resources with the first level single cycle 16K Instruction (I) and Data (D) cache that is backed by two second level caches a nine cycle 512K I cache and an eight cycle 256K D cache.

The core floor plan is optimized around the performance-sensitive single-cycle integer execution and first-level data cache (Fig. 4.8.2). Timely delivery of virtual addresses to the first level TLB require placing it within the IEU data path. Way muxing and data rotation circuitry is positioned directly below the integer vertical centerline to speed data return from the FLD to the IEU. The floor plan optimizations helped enable a 25% reduction in cycle time for the pre-silicon design target over the previous generation core.

To further improve Vmin operation the pre-silicon frequency analysis is primarily done at low voltage with heuristic runs at high voltage. Vmin sensitive circuits, including pulse latches, dynamic logic and NFET-only latches are avoided to enable robust low-voltage operation and lower power. Register files (RF) use fully interrupted feedback cells for low-voltage writes and static global bitlines where possible. (Fig. 4.8.3) A full-scan methodology supports an ATPG scan-based testing flow that enables excellent manufacturing test coverage. All RFs contain a local direct-access testing port. A fine-grain clock vernier system facilitates speed path debug through the insertion of 45000 clock skew adjustment points. The clock vernier [3] circuit has separate edge controls enabling both duty cycle adjustment and insertion delay modification to improve frequency or robustness without doing a silicon stepping. (Fig. 4.8.4)

Techniques such as eliminating domino logic in large data-paths, replacing architectural stalls with replays, eliminating glitches, and increasing dynamic clock gating efficiency to over 85% effectively reduce dynamic power in the core by an additional 60% beyond the technology scaling (Fig. 4.8.5). Post-timing analysis circuit downsizing maximizes the use of lower-leakage devices in the core (>81%) and uncore (76%), reducing overall leakage by 30%.

An improved digital activity sensor includes the ability to monitor the >30% dynamic power consumed in data patterns and reacts to power events in under 1µs. This system monitors 1834 architectural and data events to predict core power consumption. To monitor the power supply in real time, the design features an on-die droop measurement [4] with the ability to introduce controlled droop events to improve debug. The power supplies are divided into four individually regulated core pairs with additional supplies for the system interface/large caches and the I/O subsystem. The power on Poulson is distributed as follows: ~55% across the 8 cores, ~35% in the uncore and ~10% consumed in the I/O. Core pairs can be turned off for core defeatured configurations and for power savings. Individual regulation of the core pairs allows for optimization of frequency by compensating for within-die technology variations with voltage adjustments unique to each core pair. Core power supplies are bypassed by an embedded array capacitance in the central layers of the package to improve di/dt-induced noise.

The system interconnect is provided by an integrated 10-port router that connects to external IO and processors via four full-width and two half-wide point-to-point 6.4GT/s Quickpath™ (QPI) link interfaces. The ring-based system interface provides a theoretical peak bandwidth of 700GB/s (Fig. 4.8.6). The chip includes two integrated memory controllers each supporting two scalable memory interconnect (SMI) links operating in lockstep. These four SMI ports provide a 6.4GT/s connection to up to 512GB memory per socket. The SMI and QPI links provide reliability, availability, and scalability (RAS) support for features such as clock and lane failover.

In order to enable mission critical servers and the associated high RAS requirements several improvements were made to the design. These improvements include:

- Large cache arrays covered by ECC including the large L3 utilizing DECTED and protecting the MLI/MLD with inline SECDED.
- Extensive parity protection and parity interleaving on nearly all RFs.
- End-to-end parity protection with recovery-support on all critical internal buses and data paths including the ring.
- Residue protection on key logic and execution data paths.
- The adoption of radiation-hardened (RAD) sequential latching elements for vulnerable architectural and state.

The adoption of these techniques requires design methodology and automation enhancements to ensure the RAS goals are achieved without dramatically increasing die area and power. The random logic synthesis (RLS) flows optimally select hardened sequential usage based upon attack vulnerability and timing criticality. Additionally, a new error micro-architecture combines uniform logging and configuration with key hardware/firmware hooks enabling increased availability and serviceability. These techniques enabled 2× the cores and 1.5× the transistors with a lower overall susceptibility to radiation induced error events as compared to the previous generations.

Acknowledgements:
The authors thank the entire design teams from Fort Collins, CO and Hudson, MA.

References:

[1] Stackhouse, B.; et. al.; , "A 65 nm 2-Billion Transistor Quad-Core Itanium® Processor," *IEEE J. Solid-State Circuits,* 2009.
[2] Packan, P; et. al.; , "High Performance 32nm Logic Technology Featuring 2nd Generation High-k + Metal Gate Transistors," *IEDM 2009*,
[3] Mahoney, P.; Fetzer, E.; et al., "Clock distribution on a dual-core, multi-threaded Itanium®-family processor," *ISSCC Dig. Tech. Papers,* Feb. 2005.
[4] Petersen, R.; et. al.; , "Voltage transient detection and induction for debug and test," *Test Conference*, pp.1-10, Nov. 2009.

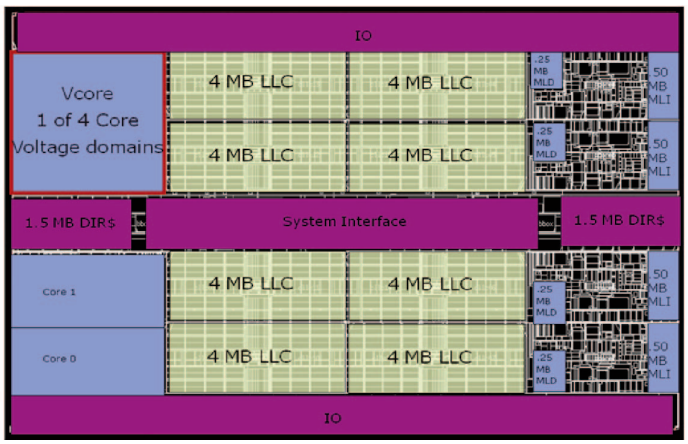

Figure 4.8.1: Poulson Processor Block Diagram.

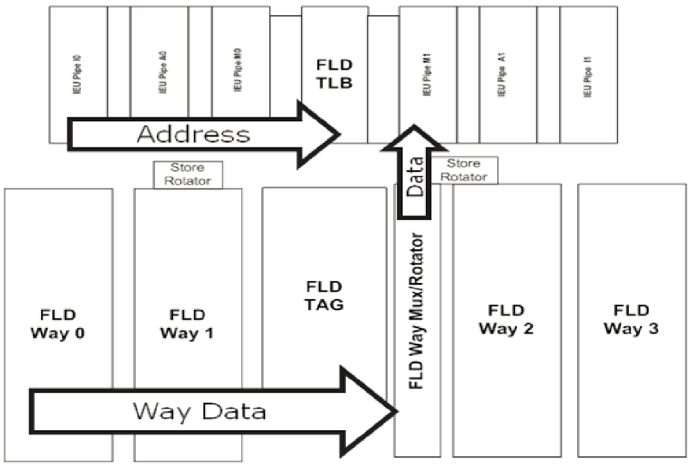

Figure 4.8.2: EXE/FLD datapath integration.

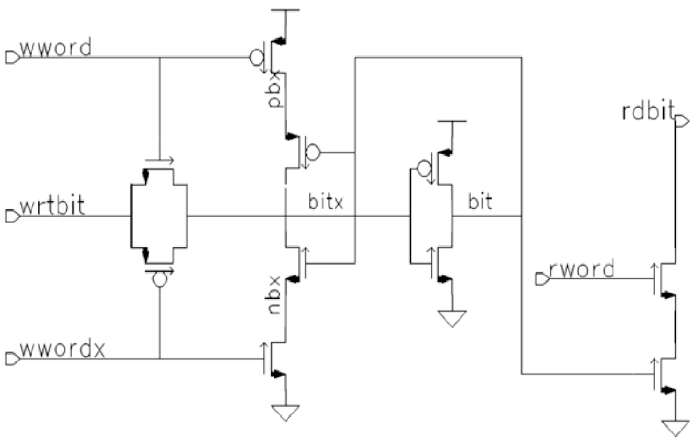

Figure 4.8.3: Fully interrupted RF bit cell.

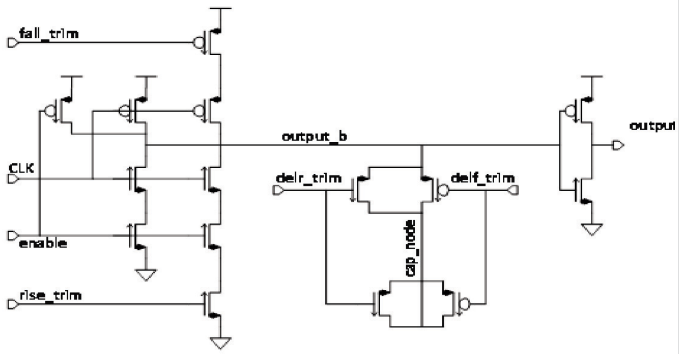

Figure 4.8.4: Clock Vernier Circuit.

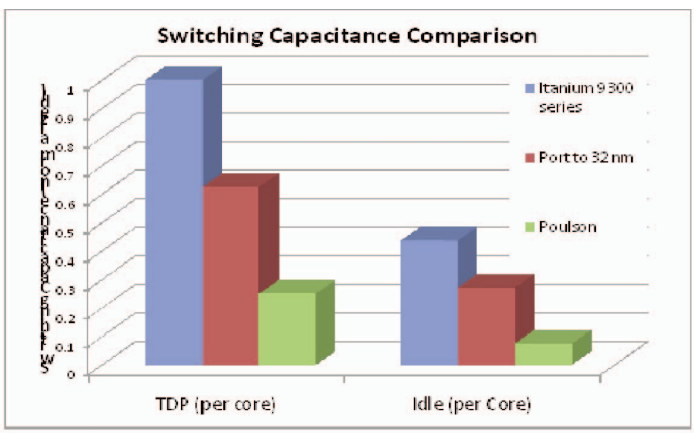

Figure 4.8.5: Power Reduction in Poulson Core.

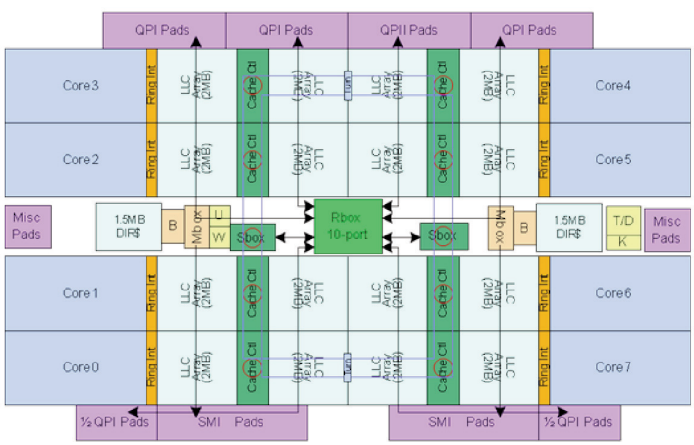

Figure 4.8.6: Poulson Block Diagram.

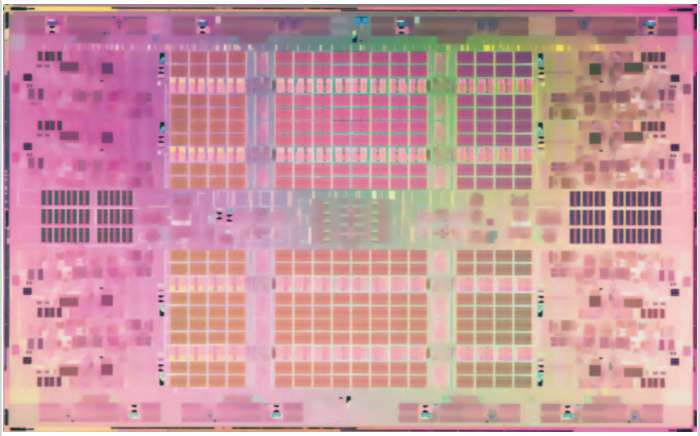

Figure 4.8.7: Poulson Die Photo.

978-1-61284-302-5/11/$26.00